文章编号: 1001-0920 (2011) 06-0851-06

不完备信息系统中的否定决策规则和知识约简

张 明[1,2], 唐振民[1], 杨习贝[1]
(1. 南京理工大学 计算机科学与技术学院，南京 210094;
2. 江苏科技大学 计算机科学与工程学院，江苏 镇江 212003)

摘 要: 为了从不完备信息系统中获得否定决策规则, 提出了描述子否定支撑集的概念, 基于此定义了下、上近似集合, 讨论其性质, 并给出了如何通过该模型获取确定性和可信性否定决策规则的方法. 为了便于应用, 给出基于分辨矩阵的否定决策规则约简的方法, 实例分析的结果表明了该方法的有效性.

关键词: 否定决策规则；描述子；粗糙集；约简
中图分类号: TP18 **文献标识码:** A

Negative decision rules and knowledge reduction in incomplete information system

ZHANG Ming[1, 2], *TANG Zhen-min*[1], *YANG Xi-bei*[1]
(1. School of Computer Science and Technology，Nanjing University of Science and Technology，Nanjing 210094, China；2. School of Computer Science and Engineering，Jiangsu University of Science and Technology，Zhenjiang 212003，China．Correspondent：ZHANG Ming，E-mail：zm_fred@163.com)

Abstract: To derive the negative rules from the incomplete decision information system, the concept of negative support set of descriptors is proposed. The lower and upper approximations based on such negative support set are presented, and their properties are discussed. Moreover, the approach to acquire the certain and credible negative rules by the proposed lower and upper approximate is presented. For facilitate application, an approach of discernibility matrix is proposed to obtain the reduction, and the results of the illustrative examples show the effectiveness of the approach.

Key words: negative decision rule；descriptor；rough sets；reduction

1 引 言

粗糙集理论[1]是一种能够定量分析和处理不精确、不一致、不完整信息与知识的数学工具. 通过下、上近似的概念, 可以在不需要任何先验知识的情况下, 从信息表中获得有价值、具有潜在效用的知识. 传统的粗集理论是建立在完备信息系统上的, 而实际要处理的信息多是不完备的. 不完备信息系统是指信息系统中出现了未知属性值, 主要分为两种情况[2]: 1) 未知属性值仅仅是被遗漏的, 但又是确实存在的 (遗漏型), 这种遗漏型未知属性值可以用任意可能的值代替; 2) 未知属性值被认为是丢失的, 是不允许被比较的 (缺席型). 处理遗漏型的情况, 主要有基于容差关系粗糙集模型[3]、限制容差关系粗糙集模型[4]等; 针对缺席型情况, 主要有非对称相似关系粗糙集模型[5]等. 这些方法总体而言是基于不可分辨关系的粗糙集模型的拓展, 通过决策类的下、上近似集可以推导出确定性和可能性两种类型的决策规则, 且这两种类型的决策规则均是肯定性质的. 即只要对象的条件属性值满足某些特定的条件, 便可以判断其决策属性值 (或将其划入某个决策类). 然而在实际应用中, 有时不能判断某个对象是否属于某个决策类别, 但是可以明确知道若对象的条件属性不满足某些特定的条件, 则可以将其从某个决策类中剔除出去 (或者说不属于某个决策类), 称此类规则为否定决策规则. 文献 [6] 较早地研究了此类否定决策规则, 并成功地将其应用于医疗专家系统中. [7] 通过研究概率粗集的正域、负域和边界域, 讨论了肯定规则、否定规则和边界规则的获取问题. [8] 给出了差异关系粗集

收稿日期: 2010-03-11；**修回日期:** 2010-05-10.
基金项目: 国家自然科学基金项目(60472060, 60572034)；国家自然科学基金重点项目(60632050).
作者简介: 张明(1978−), 男, 讲师, 博士生, 从事粗糙集及其应用的研究；唐振民(1961−), 男, 教授, 博士生导师, 从事模式识别与智能控制系统等研究.

模型, 并以此研究缺席型不完备决策信息系统中否定决策规则的获取和约简问题. 但是上述方法不适合处理遗漏型未知属性值的不完备决策信息系统.

鉴于上述分析, 本文以遗漏型未知属性值的不完备决策信息系统为研究对象, 通过考查样本的条件属性值的潜在不一致(差异)性, 提出描述子否定支撑集的概念, 用其定义否定决策类的下、上近似集以获取否定决策规则. 同时, 为了获取简化的否定决策规则, 还定义了两种可以保持下、上近似集合分布一致的相对约简, 并给出了求取这些约简的分辨矩阵方法. 实例结果表明了所提出方法的有效性.

2 基本概念

2.1 不完备决策信息系统

决策信息系统(DIS)是指既有条件属性又有决策属性的一种信息系统, 研究条件属性与决策属性间的关系问题. 一个决策信息系统为一个四元组, 即

$$\mathrm{IDIS} = \langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle.$$

其中: U 是非空有限对象集合, 称为论域; AT 是非空有限条件属性集合, $\forall a \in \mathrm{AT}$, V_a 是属性 a 的值域, d 是决策属性, 且 $\mathrm{AT}\bigcap\{d\} = \varnothing$; $V = V_{\mathrm{AT}} \bigcup V_d$ 是全体属性的值域集合; f 是信息函数. $\forall x \in U, a \in (\mathrm{AT}\bigcup\{d\})$, 定义 $f(x,a)$ 为 x 在属性 a 上的取值, 则有 $f(x,a) \in V_a$.

不完备决策信息系统(IDIS)是指信息系统中出现了未知属性值. 本文主要讨论遗漏型未知条件属性值的情况[2], 用"$*$"表示, 即 $f(x,a) = *(x \in U, a \in \mathrm{AT})$, 并假设未知属性值仅出现在条件属性值中. 一个不完备决策信息系统的实例如表 1 所示.

表 1 不完备决策信息系统

U	a	b	c	d
x_1	2	1	1	1
x_2	3	2	*	3
x_3	2	*	2	2
x_4	*	2	2	2
x_5	3	2	1	2
x_6	3	2	3	3
x_7	*	1	1	1

2.2 容差关系

定义 1 设 $\mathrm{IDIS} = \langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle$, $\forall A \subseteq \mathrm{AT}$. 由 A 决定的容差关系记为 $T(A)$, 且有

$$T(A) = \{(x,y) \in U^2 : f(x,a) = f(y,a) \bigvee f(x,a) = * \bigvee f(y,a) = *,\ \forall a \in A\}.$$

容差关系满足自反性和对称性, 而不一定满足传递性.

$\forall x \in U$, x 的容差类记为 $T_A(x)$, 且 $T_A(x) = \{y \in U : (x,y) \in T(A)\}$, 即 $T_A(x)$ 是所有与 x 具有容差关系 $T(A)$ 的对象的集合.

定义 2 设 $\mathrm{IDIS} = \langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle$, $A \subseteq \mathrm{AT}$, $\forall X \in U$, X 基于 $T(A)$ 的下、上近似集分别记为 $\underline{A}_T(X)$ 和 $\overline{A}_T(X)$, 且有

$$\underline{A}_T(X) = \{x \in U : T_A(X) \subseteq X\}, \tag{1}$$

$$\overline{A}_T(X) = \{x \in U : T_A(X) \bigcap X \neq \varnothing\}. \tag{2}$$

2.3 决策规则获取

在粗集理论中, 隐藏在信息系统中的知识通常以决策规则的形式被挖掘出来, 考察对象(或训练样本) $x \in U$ 得到如下决策规则:

$$r_x:\ t \to s. \tag{3}$$

其中: t 是该决策规则的条件部分, $t = \bigwedge(a, v_a), a \in A \subseteq \mathrm{AT}$; s 是决策部分, $s = (d,i), i \in V_d$, i 是类别标签. 此种决策规则可以描述为: 若对象的条件属性满足某个特定的值, 则其决策属性一定可以取得某个特定的值. 文献[8-10]研究了此类决策规则的获取和约简问题.

本文认为, 现实世界中除了上述肯定性质的决策规则外, 还应该有与之对应的否定性质的决策规则来帮助决策者作出否定判决, 即若考查对象的条件属性不满足某些特定的条件, 则其决策属性不可能获得某些特定的取值(或从某个决策类中剔除). 相对于上述肯定性质的决策规则的一般形式, 可以给出否定决策规则的一般形式为

$$r_x:\ \neg t \to \neg s. \tag{4}$$

其中: $\neg t$ 是否定决策规则的条件部分, $t = \bigwedge(a, v_a), a \in A \subseteq \mathrm{AT}$; $\neg s$ 是决策部分, $s = (d,i), i \in V_d$. 上述否定决策规则中的 $t = \bigwedge(a, v_a)$ 恰好是 Leung 等人[9]定义的描述子. 为了便于表述如何从不完备决策信息系统中挖掘出否定决策规则, 下文通过考查对象间的条件属性的差异性, 提出一个描述子的否定支撑集的概念.

3 描述子的否定支撑集与粗糙集

定义 3 设 $\mathrm{IDIS} = \langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle$, $\forall a \in \mathrm{AT}$, $v \in V_{\mathrm{AT}}$, 称 (a,v) 为一个 AT- 原子公式. 任意一个或几个不同的 AT- 原子公式的合取称为一个 AT- 条件描述子, 记为 $t = \bigwedge_{a \in A}(a, v_a), A \subseteq \mathrm{AT}$. 若 t 为 AT- 条件描述子, t 中出现的所有属性的集合记为 $\mathrm{AT}(t)$, $\|t\| = \{x \in U : f(x,a) = * \bigvee f(x,a) \neq v_a, \forall a \in \mathrm{AT}(t)\}$, 则称 $\|t\|$ 为描述子 t 的否定支撑集. 在表 1 的不完备决策信息系统中, 令 $t = (a,2)\bigwedge(b,1)\bigwedge(c,1)$, 有 $\|t\| = \{x_2, x_4, x_6\}$, 则 $\forall x \in \|t\|$, 一定有 $f(x,a) \neq 2$, $f(x,b) \neq 1$, 且 $f(x,c) \neq 1$. 设 $V_d = \{1, 2, \cdots, i\}$, 若 $s = (d,i)$, 则称 s 为决策描述子. 令 $\|s\|$ 为决策描述子 s 的否定支撑集, 有 $\|s\| = \{x \in U : f(x,d) \neq i\}$. 若 t 和 t' 是两个 AT- 条件描述子, $\forall (a,v) \in t$, 均有 $(a,v) \in t'$, 则

称 t 比 t' 更为粗或者 t' 比 t 更为细, 记为 $t \succeq t'$ 或 $t' \preceq t$; 若 t 由 t' 中出现的原子公式的真子集构成, 则称 t 比 t' 严格粗或者 t' 比 t 严格细, 记为 $t \succ t'$ 或 $t' \prec t$.

注 1　根据条件描述子否定支持集的定义, 设 t 为一个 AT- 条件描述子, 且 $x \in \|t\|$. 若 $\exists a \in \mathrm{AT}$ 使得 $f(x,a) \neq *$(此时 x 至少有一个条件属性值不为空), 则 x 必定有一个属性值与描述子 t 中的一个原子公式的属性值不一致, 这说明 x 一定不满足 t 的原子公式; 若 $\forall a \in \mathrm{AT}$ 使得 $f(x,a) = *$(此时 x 的所有条件属性值均为空), 根据未知属性值的解释[2], 则 "$*$"可能是值域 V_a 上的任意一个属性值, 因此, x 的条件属性值与描述子 t 中的原子公式的属性值有 $1/(|V_{a1}| \cdot |V_{a1}| \cdot \cdots \cdot |V_{ai}|)$ 的机会取值相同(设 $\mathrm{AT} = \{a_1, a_2, \cdots, a_i\}$, $|V_{ai}|$ 表示 V_{ai} 的基). 显然在实际应用中, 如果对象的所有条件属性值均未知, 则讨论其决策属性是没有任何意义的. 综上, 称条件描述子 t 的否定支持集是所有不满足 t 中原子公式的对象的集合.

定义 4　令 $\mathrm{IDIS} = \langle U, \mathrm{AT} \bigcup \{d\}, V, f\rangle$ 为一个不完备决策信息系统, 记 $\mathrm{DES(AT)} = \{t:\ t$ 是 AT- 条件描述子, $\|t\| \neq \varnothing\}$; 对于 $\forall t \in \mathrm{DES(AT)}$, 若 $\mathrm{AT}(t) = \mathrm{AT}$, 则称 t 为一个完全 AT- 条件描述子, 记 $\mathrm{FDES(AT)} = \{t : t \in \mathrm{DES(AT)}, t$ 是完全 AT- 条件描述子$\}$.

定义 5　令 $\mathrm{IDIS} = \langle U, \mathrm{AT} \bigcup \{d\}, V, f\rangle$, $A \subseteq \mathrm{AT}$, $\forall X \subseteq U$, X 基于属性 A 的下近似集、上近似集分别为

$$\underline{A}_{\mathrm{des}}(X) = \{\|t\| : \|t\| \subseteq X, t \in \mathrm{FDES}(A)\}, \tag{5}$$

$$\overline{A}_{\mathrm{des}}(X) = \{\|t\| : \|t\| \bigcap X \neq \varnothing,\ t \in \mathrm{FDES}(A)\}. \tag{6}$$

根据 X 的下、上近似集, 可得到其正域、边界域和负域分别为

$$\mathrm{POS}_{\mathrm{des}}(X) = \underline{A}_{\mathrm{des}}(X), \tag{7}$$

$$\mathrm{BND}_{\mathrm{des}}(X) = \overline{A}_{\mathrm{des}}(X) - \underline{A}_{\mathrm{des}}(X), \tag{8}$$

$$\mathrm{NEG}_{\mathrm{des}}(X) = \{\|t\| : t \in \mathrm{FDES}(A)\} - \underline{A}_{\mathrm{des}}(X). \tag{9}$$

推论 1　设 $\mathrm{IDIS} = \langle U, \mathrm{AT} \bigcup \{d\}, V, f\rangle$, t 为条件描述子, 对于任意 $X \subseteq U$, 有:

1) 若 $\|t\| \in \mathrm{POS}_{\mathrm{des}}(X)$, 则对于任意 $x \in \|t\|$ 一定有 $x \in X$;

2) 若 $\|t\| \in \mathrm{BND}_{\mathrm{des}}(X)$, 则对于任意 $x \in \|t\|$ 可能有 $x \in X$, 也可能有 $x \notin X$;

3) 若 $\|t\| \in \mathrm{NED}_{\mathrm{des}}(X)$, 则对于任意 $x \in \|t\|$ 一定有 $x \notin X$.

定理 1　令 $\mathrm{IDIS} = \langle U, \mathrm{AT} \bigcup \{d\}, V, f\rangle$, $A \subseteq \mathrm{AT}$, $\forall X, Y \subseteq U$, 则有以下公式成立:

$$\begin{aligned}&\underline{A}_{\mathrm{des}}(\varnothing) = \overline{A}_{\mathrm{des}}(\varnothing) = \varnothing,\\ &\underline{A}_{\mathrm{des}}(X) \subseteq \overline{A}_{\mathrm{des}}(X);\end{aligned} \tag{10}$$

$$\begin{aligned}&X \subseteq Y \Rightarrow \underline{A}_{\mathrm{des}}(X) \subseteq \underline{A}_{\mathrm{des}}(Y),\\ &\qquad \overline{A}_{\mathrm{des}}(X) \subseteq \overline{A}_{\mathrm{des}}(Y);\end{aligned} \tag{11}$$

$$\begin{aligned}&\underline{A}_{\mathrm{des}}(X \bigcap Y) = \underline{A}_{\mathrm{des}}(X) \bigcap \underline{A}_{\mathrm{des}}(Y),\\ &\overline{A}_{\mathrm{des}}(X \bigcap Y) = \overline{A}_{\mathrm{des}}(X) \bigcap \overline{A}_{\mathrm{des}}(Y);\end{aligned} \tag{12}$$

$$\begin{aligned}&\underline{A}_{\mathrm{des}}(X) \bigcup \underline{A}_{\mathrm{des}}(Y) \subseteq \underline{A}_{\mathrm{des}}(X \bigcup Y),\\ &\overline{A}_{\mathrm{des}}(X) \bigcup \overline{A}_{\mathrm{des}}(Y) \subseteq \overline{A}_{\mathrm{des}}(X \bigcup Y).\end{aligned} \tag{13}$$

证明　1) 由定义 5 易证式 (10) 成立.

2) 设任意 $\|t\| \in \underline{A}_{\mathrm{des}}(X)$, 有 $\|t\| \neq \varnothing$ 且 $\|t\| \subseteq X$. 又 $X \subseteq Y$, $\|t\| \subseteq Y$, 所以有 $\underline{A}_{\mathrm{des}}(X) \subseteq \underline{A}_{\mathrm{des}}(Y)$. 同理可得 $\overline{A}_{\mathrm{des}}(X) \subseteq \overline{A}_{\mathrm{des}}(Y)$ 也成立,于是式 (11) 成立.

3) 设任意 $\|t\| \in \underline{A}_{\mathrm{des}}(X \bigcap Y)$, 有 $\|t\| \neq \varnothing$, 且有 $\|t\| \subseteq (X \bigcap Y)$, 即 $\|t\| \subseteq X$ 且 $\|t\| \subseteq Y$, 所以有 $\|t\| \subseteq \underline{A}_{\mathrm{des}}(X)$, $\|t\| \subseteq \underline{A}_{\mathrm{des}}(Y)$, 于是有 $\|t\| \subseteq (\underline{A}_{\mathrm{des}}(X) \bigcap \underline{A}_{\mathrm{des}}(Y))$ 成立. 设任意 $\|t\| \subseteq (\underline{A}_{\mathrm{des}}(X) \bigcap \underline{A}_{\mathrm{des}}(Y))$, 则有 $\|t\| \neq \varnothing$, 且 $\|t\| \subseteq X$, $\|t\| \subseteq Y$, 即 $\|t\| \subseteq (X \bigcap Y)$, 所以 $\|t\| \in \underline{A}_{\mathrm{des}}(X \bigcap Y)$, $\underline{A}_{\mathrm{des}}(X \bigcap Y) = \underline{A}_{\mathrm{des}}(X) \bigcap \underline{A}_{\mathrm{des}}(Y)$ 成立. 同理可证 $\overline{A}_{\mathrm{des}}(X \bigcap Y) = \overline{A}_{\mathrm{des}}(X) \bigcap \overline{A}_{\mathrm{des}}(Y)$ 也成立,于是式 (12) 成立.

4) 因为 $X, Y \subseteq X \bigcup Y$, 由式 (11) 的结论可得

$$\underline{A}_{\mathrm{des}}(X) \subseteq \underline{A}_{\mathrm{des}}(X \bigcup Y),\ \underline{A}_{\mathrm{des}}(Y) \subseteq \underline{A}_{\mathrm{des}}(X \bigcup Y).$$

所以有 $\underline{A}_{\mathrm{des}}(X) \bigcup \underline{A}_{\mathrm{des}}(Y) \subseteq \underline{A}_{\mathrm{des}}(X \bigcup Y)$ 成立, 同理可得 $\overline{A}_{\mathrm{des}}(X) \bigcup \overline{A}_{\mathrm{des}}(Y) \subseteq \overline{A}_{\mathrm{des}}(X \bigcup Y)$ 也成立, 因此式 (13) 成立. □

4　否定决策规则

为了获取第 2.3 节所示的否定决策规则, 首先引入决策类补集的概念. 在不完备决策信息系统 IDIS 中, 设 $V_d = \{1, 2, \cdots, i\}$, 根据决策属性可以得到论域上 U 的一个划分 $U/d = \{X_1, X_2, \cdots, X_i\}$, 且有 $X_i = \{x \in U : f(x,d) = i\}$. 令 $\neg X_i = U - X_i$, 则 $\neg X_i$ 是决策类 X_i 的补集, 表示所有决策属性取值不为 i 的元素的集合. 令 $U/\neg d = \{\neg X_1, \neg X_2, \cdots, \neg X_i\}$, 则 $U/\neg d$ 是论域 U 所有决策类补的集合. 又根据条件描述子 t 的否定支持集的概念, $\|t\|$ 是所有不满足 t 中原子公式的对象的集合. 有了上述两个概念, 下面来研究如何获取否定决策规则.

定义 6　令 $\mathrm{IDIS} = \langle U, \mathrm{AT} \bigcup \{d\}, V, f\rangle$, 设 $V_d = \{1, 2, \cdots, i\}$, $\mathrm{AT} = \{a_1, a_2, \cdots, a_i\}$, 对于任意 $t \in \mathrm{FDES(AT)}$, t 所对应的否定决策规则 $r_t : \neg t \to \neg s$ 可以描述为

$$\begin{aligned}&(f(x,a_1) \neq v_1) \bigwedge (f(x,a_2) \neq v_2) \bigwedge \cdots \bigwedge\\ &(f(x,a_m) \neq v_m) \to f(x,d) \neq i.\end{aligned}$$

对于否定决策规则 r_t 的可信性程度由下面定义的可信性因子度量:

$$\mathrm{Cer}(r_t)=\frac{\mathrm{card}(\|t\|\bigcap\|s\|)}{\mathrm{card}(\|t\|)}.\tag{14}$$

其中: $\|t\|$ 是条件描述子的否定支撑集, $\|s\|$ 是决策描述子的否定支撑集, $\mathrm{card}(X)$ 是集合 X 的基数. 令 $\mathrm{Cer}(r_t)=\alpha$, 当 $\alpha=1$ 时, r_t 是一条确定性的否定决策规则; 当 $0<\alpha<1$ 时, r_t 是一条可信性的否定决策规则, 其可信性程度由 $100\,\alpha\%$ 来度量.

定理 2 设 $\mathrm{IDIS}=\langle U,\mathrm{AT}\bigcup\{d\},V,f\rangle$, 令 $V_d=\{1,2,\cdots,i\}$, $U/\neg d=\{\neg X_1,\neg X_2,\cdots,\neg X_i\}$, 条件描述子 $t\in\mathrm{DES(AT)}$, 决策描述子 $s=(d,i)$, 则有:

1) 若 $\|t\|\in\mathrm{POS_{des}}(\neg X_i)$, 则 r_t 是一条确定性的否定决策规则;

2) $\|t\|\in\mathrm{BND_{des}}(\neg X_i)$, 则 r_t 是一条可信性的否定决策规则.

证明 由 $\|t\|\in\mathrm{POS_{des}}(\neg X_i)$ 可知, $\|t\|\neq\varnothing$ 且 $\|t\|\subseteq\neg X_i$, $\forall x\in\|t\|$ 有 $x\in\neg X_i$, 即有 $f(x,d)\neq i$. 根据 $\|s\|$ 的定义有 $\|t\|\subseteq\|s\|$, $\mathrm{Cer}(r_t)=1$ 成立, 所以 r_t 是一条确定性的否定决策规则. 同理可证 2) 也成立. □

例 1 求表 1 所示的不完备决策信息系统中所有确定性和可信性否定决策规则. 易知

$$U=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\},\ \mathrm{AT}=\{a,b,c\},$$
$$V_a=V_b=V_c=V_d=\{1,2,3\}.$$

由定义 3 可知

$$t_1=(a,1)\bigwedge(b,1)\bigwedge(c,1),\ \|t_1\|=\{x_2,x_3,x_4,x_6\};$$
$$t_2=(a,1)\bigwedge(b,2)\bigwedge(c,2),\ \|t_2\|=\{x_1,x_7\};$$
$$t_3=(a,2)\bigwedge(b,1)\bigwedge(c,1),\ \|t_3\|=\{x_2,x_4,x_6\};$$
$$t_4=(a,2)\bigwedge(b,1)\bigwedge(c,2),\ \|t_4\|=\{x_2,x_5,x_6\};$$
$$t_5=(a,2)\bigwedge(b,2)\bigwedge(c,2),\ \|t_5\|=\{x_7\};$$
$$t_6=(a,2)\bigwedge(b,3)\bigwedge(c,2),\ \|t_6\|=\{x_2,x_5,x_6,x_7\};$$
$$t_7=(a,3)\bigwedge(b,1)\bigwedge(c,1),\ \|t_7\|=\{x_3,x_4\};$$
$$t_8=(a,3)\bigwedge(b,2)\bigwedge(c,1),\ \|t_8\|=\{x_3\};$$
$$t_9=(a,3)\bigwedge(b,2)\bigwedge(c,2),\ \|t_9\|=\{x_1,x_7\};$$
$$t_{10}=(a,3)\bigwedge(b,2)\bigwedge(c,3),\ \|t_{10}\|=\{x_1,x_3,x_7\}.$$

论域 U 上所有决策类的补集满足

$$U/\neg d=\{\neg X_1,\neg X_2,\neg X_3\}=$$
$$\{\{x_2,x_3,x_4,x_5,x_6\},\{x_1,x_2,x_6,x_7\},$$
$$\{x_1,x_3,x_4,x_5,x_7\}\}.$$

由定义 5 可知

$$\mathrm{POS_{des}}(\neg X_1)=\{\|t_1\|,\|t_3\|,\|t_4\|,\|t_7\|,\|t_8\|\},$$
$$\mathrm{POS_{des}}(\neg X_2)=\{\|t_2\|,\|t_5\|,\|t_9\|\},$$
$$\mathrm{POS_{des}}(\neg X_3)=\{\|t_2\|,\|t_5\|,\|t_7\|,\|t_8\|,\|t_9\|,\|t_{10}\|\};$$
$$\mathrm{BND_{des}}(\neg X_1)=\{\|t_6\|,\|t_{10}\|\},$$
$$\mathrm{BND_{des}}(\neg X_2)=\{\|t_1\|,\|t_3\|,\|t_4\|,\|t_6\|,\|t_{10}\|\},$$
$$\mathrm{BND_{des}}(\neg X_3)=\{\|t_1\|,\|t_3\|,\|t_4\|,\|t_6\|\}.$$

根据定理 2 可得确定性的否定决策规则为

$$f(x,a)\neq1\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq1\rightarrow f(x,d)\neq1,$$
$$f(x,a)\neq1\bigwedge f(x,b)\neq2\bigwedge f(x,c)\neq2\rightarrow$$
$$f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq2\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq1\rightarrow f(x,d)\neq1,$$
$$f(x,a)\neq2\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq2\rightarrow f(x,d)\neq1,$$
$$f(x,a)\neq2\bigwedge f(x,b)\neq2\bigwedge f(x,c)\neq2\rightarrow$$
$$f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq3\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq1\rightarrow$$
$$f(x,d)\neq1\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq3\bigwedge f(x,b)\neq2\bigwedge f(x,c)\neq1\rightarrow$$
$$f(x,d)\neq1\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq3\bigwedge f(x,b)\neq2\bigwedge f(x,c)\neq2\rightarrow$$
$$f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq3\bigwedge f(x,b)\neq2\bigwedge f(x,c)\neq3\rightarrow f(x,d)\neq3.$$

根据定理 2 可得到可信性的否定决策规则为

$$f(x,a)\neq1\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq1\rightarrow$$
$$f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq2\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq1\rightarrow$$
$$f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq2\bigwedge f(x,b)\neq1\bigwedge f(x,c)\neq2\rightarrow$$
$$f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq2\bigwedge f(x,b)\neq3\bigwedge f(x,c)\neq2\rightarrow$$
$$f(x,d)\neq1\bigwedge f(x,d)\neq2\bigwedge f(x,d)\neq3,$$
$$f(x,a)\neq3\bigwedge f(x,b)\neq2\bigwedge f(x,c)\neq3\rightarrow$$
$$f(x,d)\neq1\bigwedge f(x,d)\neq2.$$

5 知识约简

知识约简[6,8-10]是粗糙集理论研究的核心问题, 借助知识约简能剔除决策规则中的冗余属性, 从而得到简化的决策规则. 为了对本文所讨论的否定决策规则进行约简, 下面将分别给出保持下、上近似分布一致的相对约简方法.

定义 7 设 $\mathrm{IDIS}=\langle\mathrm{AT}\bigcup\{d\},V,f\rangle$, $V_d=\{1,2,\cdots,i\}$, $U/\neg d=\{\neg X_1,\neg X_2,\cdots,\neg X_i\}$, 条件描述子 t, $t'\in\mathrm{DES(AT)}$, 且 $t'\geqslant t$, 令

$$L(t)=\{\neg X_i:\|t\|\in\underline{\mathrm{AT}}_{\mathrm{des}}(\neg X_i),\neg X_i\in U/\neg d\},$$
$$H(t)=\{\neg X_i:\|t\|\in\overline{\mathrm{AT}}_{\mathrm{des}}(\neg X_i),\neg X_i\in U/\neg d\}.$$

1) 若 $L(t) = L(t')$, 且 $\forall t'' \succeq t$, $L(t) \neq L(t'')$, 则称 t' 是 t 的一个下近似分布一致的相对约简属性描述子.

2) 若 $H(t) = H(t')$, 且 $\forall t'' \succeq t$, $H(t) \neq H(t'')$, 则称 t' 是 t 的一个上近似分布一致的相对约简属性描述子.

注 2　从定义 7 可以看出, t 的下近似 (上近似) 相对约简描述子 t' 同时满足 $\|t'\|$ 是 $\neg X_i$ 的下近似 (上近似), 且 t' 由 t 中最小数量的原子公式合取构成, t 的所有下近似 (上近似) 相对约简描述子的条件属性集记为 $\mathrm{red}_L(t)(\mathrm{red}_H(t))$.

定义 8　设 IDIS=$\langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle$, V_d=$\{1, 2, \cdots, i\}$, $U/\neg d$=$\{\neg X_1, \neg X_2, \cdots, \neg X_i\}$, t∈FDES(AT). 记 $D_L(t) = \{x : \|t\| \in \underline{\mathrm{AT}}_{\mathrm{des}}(\neg X_i), x \in X_i\}$, $D_H(t) = \{x : \|t\| \bigcap \neg X_i = \varnothing, x \in X_i\}$. 定义 $D_L(t, x)$ 为不完备决策信息系统中条件描述子 t 的下近似可辨识条件属性集, $D_H(t, x)$ 为上近似可辨识条件属性集, 且有

$$D_L(t,x) = \begin{cases} \{a \in \mathrm{AT} : x \notin \|(a, a_v)\|\},\ x \in D_L(t); \\ \mathrm{AT},\ x \notin D_L(t). \end{cases}$$

$$D_H(t,x) = \begin{cases} \{a \in \mathrm{AT} : x \notin \|(a, a_v)\|\},\ x \in D_H(t); \\ \mathrm{AT},\ x \notin D_H(t). \end{cases}$$

定理 3　设 IDIS $= \langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle$, 若属性描述子 $t \in \mathrm{FDES(AT)}$, $t' \in \mathrm{DES(AT)}$, 且 $t' \geqslant t$, 则有

$$L(t) = L(t') \Leftrightarrow D_L(t,x) \bigcap \mathrm{AT}(t') \neq \varnothing,\quad \forall x \in D_L(t); \tag{15}$$

$$H(t) = H(t') \Leftrightarrow D_H(t,x) \bigcap \mathrm{AT}(t') \neq \varnothing,\quad \forall x \in D_H(t). \tag{16}$$

证明　1) 充分性. 根据定义 8, 假设 $\exists x \in D_L(t)$ (即 $x \in X_i$ 且 $\|t\| \in \neg X_i$) 和 $x \notin \|t\|$ 使得 $D_L(t, x) \bigcap \mathrm{AT}(t') = \varnothing$ 成立, 因为 $t' \geqslant t$, 所以有 $x \in \|t'\|$. 根据条件有 $L(t) = L(t')$, 即 $\|t\| \subseteq \neg X_i \Leftrightarrow \|t'\| \subseteq \neg X_i$, 此时有 $x \in \neg X_i$, 这与 $x \in X_i$ 相矛盾.

2) 必要性. 由 $t' \geqslant t$, 可得 $L(t) \supseteq L(t')$, 故只需证明 $L(t) \subseteq L(t')$. 对于 $L(t) \neq \varnothing$, $\neg X_i \in L(t)$(即 $\|t\| \subseteq \neg X_i$) 有: 对于 $\forall x \in X_i$, $D_L(t, x) \bigcap \mathrm{AT}(t') \neq \varnothing$ 成立, 即 $x \notin \|t'\|$. 因为 x 是 X_i 中任意选定的, 所以此时必定有 $\|t'\| \subseteq \neg X_i$, 从而 $L(t) \subseteq L(t')$. 综上可得式 (15) 成立, 同理可以证明式 (16) 也成立. □

定义 9　设 IDIS $= \langle U, \mathrm{AT}\bigcup\{d\}, V, f\rangle$, 条件描述子 $t \in \mathrm{FDES(AT)}$, 令

$$\Delta_L(t) = \bigwedge_{X \in D_L(t)} \left(\bigvee D_L(t, x)\right);$$

$$\Delta_H(t) = \bigwedge_{X \in D_H(t)} \left(\bigvee D_H(t, x)\right).$$

称 $\Delta_L(t)$ 和 $\Delta_H(t)$ 分别为描述子 t 的下近似和上近似分布一致相对约简区分函数.

根据布尔推理理论, 由定义 9 和定理 3 可得如下结论.

定理 4　$\Delta_L(t)$ 和 $\Delta_H(t)$ 的极小析取范式中的每个合取项所对应的属性子集即为 t 的下近似和上近似分布一致的相对约简描述子中出现的属性集合.

例 2　求解例 1 中确定性否定决策规则的约简.

Step 1: 根据定义 8 可得描述子 t 的可辨识条件属性集 $D(t, x)$ 如表 2 所示.

表 2　描述子t的可辨识条件属性集

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
t_1	bc	AT	AT	AT	AT	AT	bc
t_2	AT	b	c	bc	b	b	AT
t_3	bc	AT	AT	AT	AT	AT	bc
t_4	ab	AT	AT	AT	AT	AT	b
t_5	AT	b	ac	bc	b	b	AT
t_7	bc	a	AT	AT	AT	a	bc
t_8	c	ab	AT	AT	AT	ab	c
t_9	AT	ab	c	bc	ab	ab	AT
t_{10}	AT	ab	AT	AT	AT	abc	AT

Step 2: 根据定义 9 和定理 4 可以得到描述子 t 的所有相对约简的条件属性集如下:

$\mathrm{red}_L(t_1) = \Delta_L(t_1) = (b \bigvee c) \bigwedge (b \bigvee c) = b \bigvee c$,

$\mathrm{red}_L(t_2) = b \bigwedge c \bigwedge (b \bigvee c) \bigwedge b \bigwedge b = b \bigwedge c$,

$\mathrm{red}_L(t_3) = b \bigvee c$, $\mathrm{red}_L(t_4) = b$,

$\mathrm{red}_L(t_5) = (a \bigwedge b) \bigvee (b \bigwedge c)$,

$\mathrm{red}_L(t_7) = (a \bigwedge b) \bigvee (a \bigwedge c)$,

$\mathrm{red}_L(t_8) = (a \bigwedge c) \bigvee (b \bigwedge c)$,

$\mathrm{red}_L(t_9) = (a \bigwedge c) \bigvee (b \bigwedge c)$,

$\mathrm{red}_L(t_{10}) = a \bigvee b$.

Step 3: 根据 t 的相对约简条件属性集对确定性的否定决策规则进行约简, 有

$f(x, b) \neq 1 \to f(x, d) \neq 1$, $f(x, c) \neq 1 \to f(x, d) \neq 1$,

$f(x, a) \neq 3 \to f(x, d) \neq 3$, $f(x, b) \neq 2 \to f(x, d) \neq 3$,

$f(x, b) \neq 2 \bigwedge f(x, c) \neq 2 \to f(x, d) \neq 2$,

$f(x, a) \neq 2 \bigwedge f(x, b) \neq 2 \to f(x, d) \neq 2$,

$f(x, a) \neq 3 \bigwedge f(x, b) \neq 1 \to f(x, d) \neq 1$,

$f(x, a) \neq 3 \bigwedge f(x, c) \neq 1 \to f(x, d) \neq 1$,

$f(x, b) \neq 2 \bigwedge f(x, c) \neq 1 \to f(x, d) \neq 1$,

$f(x, a) \neq 3 \bigwedge f(x, b) \neq 2 \to f(x, d) \neq 2$.

6　结　论

利用粗糙集理论在不完备决策信息系统中获取决策规则一直是粗糙集理论研究的热点和难点问题.

本文从对象属性间的差异性出发, 定义了描述子否定支持集的概念, 并据此给出了否定决策规则获取和约简的方法. 该否定决策规则不仅可以应用于专家系统的决策分析中[6], 而且可以应用于人工智能的规则搜索和模式分类中, 这将是下一步的研究工作.

参考文献(References)

[1] Pawlak Z. Rough sets[J]. Int J of Computer and Information Sciences, 1984, 11(5): 341-356.

[2] Grzymala-Busse J W. Data with missing attribute values: Generalization of indiscernibility relation and rule reduction[M]. Berlin: Springer-Verlag, 2004.

[3] Kryszkiewicz M. Rough set approach to incomplete information systems[J]. Information Sciences, 1998, 112(1/2/3/4): 39-49.

[4] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243.
(Wang G X. Extension of rough set under incomplete information systems[J]. J of Computer and Development, 2002, 39(10): 1238-1243.)

[5] Stefanowski J, Tsoukias A. Incomplete information tables and rough classification[J]. Computational Intelligence, 2001, 17(3): 545-566.

[6] Tsumoto S. Automated discovery of positive and negative knowledge in clinical databases[J]. Engineering in Medicine Biology, 2000, 19(4): 56-62.

[7] Yao Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180(3): 341-353.

[8] Yang X B, Yu D J, Yang J Y, et al. Difference relation-based rough set and negative rules in incomplete information system[J]. Int J of Uncertainty, Fuzziness and Knowledge-based Systems, 2009, 17(5): 649-665.

[9] Leung Y, Wu W Z, Zhang W X. Knowledge acquisition in incomplete information systems: A rough set approach[J]. European J of Operational Research, 2006, 168(1): 164-180.

[10] 黄兵, 胡作进, 周献中. 优势模糊粗糙模型及其在审计风险评估中的应用[J]. 控制与决策, 2009, 24(6): 899-902.
(Huang B, HU Z J, Zhou X Z. Dominance relation-based fuzzy-rough model and its application to auit risk evaluation[J]. Control and Decision, 2009, 24(6): 899-902.)

(上接第850页)

[4] Torra V, Miyamoto S, Endo Y, et al. On intuitionistic fuzzy clustering for its application to privacy[C]. Proc of the 2008 IEEE/FUZZ Int Conf on Fuzzy Systems. Hong Kong: IEEE Press, 2008: 1042-1048.

[5] Iakovidis D K, Pelekis N, Kotsifakos E, et al. Intuitionistic fuzzy clustering with applications in computer vision[J]. Lecture Notes in Computer Science, 2008, 5259(1): 764-774.

[6] 吴成茂. 模糊 C-均值算法在直觉模糊数聚类中的应用[J]. 计算机工程与应用, 2009, 45(16): 141-145.
(Wu C M. Fuzzy C-means algorithm applied in intuitionistic fuzzy numbers clustering[J]. Computer Engineering and Applications, 2009, 45(20): 141-145.)

[7] 徐小来, 雷英杰, 赵学军. 基于直觉模糊熵的直觉模糊聚类[J]. 空军工程大学学报, 2008, 9(2): 80-83.
(Xu X L, Lei Y J, Zhao X J. Intuitionistic fuzzy clustering based on intuitionistic fuzzy entropy[J]. J of Air Force Engineering University, 2008, 9(2): 80-83.)

[8] 申晓勇, 雷英杰, 蔡茹, 等. 一种基于密度函数的直觉模糊聚类初始化方法[J]. 计算机科学, 2009, 36(5): 197-199.
(Shen X Y, Lei Y J, Cai R, et al. Initialization method for intuitionistic fuzzy clustering based on density function[J]. Computer Science, 2009, 36(5): 197-199.)

[9] 申晓勇, 雷英杰, 蔡茹, 等. 基于目标函数的直觉模糊集合数据的聚类方法[J]. 系统工程与电子技术, 2009, 31(7): 2732-2735.
(Shen X Y, Lei Y J, Cai R, et al. Clustering technique to intuitionistic fuzzy sets data based on objective function[J]. Systems Engineering and Electronics, 2009, 31(11): 2732-2735.)

[10] Wang W Q, Xin X L. Distance measure between intuitionistic fuzzy sets[J]. Pattern Recognition Letters, 2005, 26(3): 2063-2069.

[11] Li D F. Some measures of dissimilarity in intuitionistic fuzzy structures[J]. J of Computer and System Sciences, 2004, 68(1): 115-122.

[12] Liu H W. New similarity measures between intuitionistic fuzzy sets and between elements[J]. Mathematical and Computer Modelling, 2005, 42(1): 61-70.